An Enhanced System for Link and Mode Identification for GPS-based Personal Travel Surveys

Sheung Yuen Amy Tsui
Department of Civil Engineering
University of Toronto, 35 St. George Street, Toronto, Ontario, Canada M5S 1A4
Telephone: 1-416-978-5049
Email: amysy.tsui@utoronto.ca

Amer S. Shalaby, Ph.D., P.Eng.
Assistant Professor, Department of Civil Engineering
University of Toronto, 35 St. George Street, Toronto, Ontario, Canada M5S 1A4
Telephone: 1-416-978-5907
Fax: 1-416-978-5054
Email: amer@ecf.utoronto.ca

Word Count:

- Text	= 6,517
- Tables (3 x 250)	= 750
- Figure (1 x 250)	= 250
	= 7,517

Date of Submission: **November 13, 2005**

ABSTRACT

This project developed an integrated GPS-GIS system to automate the processing of GPS-based personal travel survey data. Two versions of the analysis system were developed in this project: a GPS-Alone System, which only uses GPS travel data as input, and a GPS/GIS Integrated System, which uses both GPS travel data and topological information in a GIS platform as input. The GPS-Alone System includes an activity identification algorithm and a fuzzy logic-based mode identification algorithm. The GPS/GIS Integrated System includes link identification in a GIS platform, as well as an Interactive Link Matching-Mode Identification Subsystem, which further refines the results from previous identification performed separately. This project demonstrates how GPS travel data analysis can be automated and highlights the benefits brought by an interactive analysis system, providing an innovative analysis method for personal-based GPS multimodal travel surveys.

INTRODUCTION

The quality and scope of transportation models are highly dependent on those of the used travel data, which are predominantly collected through travel surveys. Traditionally, travel surveys have been conducted through mail and telephone interviews. In the last decade, computer-assisted telephone interview (CATI) and the Internet have also been utilized in travel surveys. However, these survey methods have several known limitations, including under-reporting of trips and the limited level of detail of survey data that can practically and reliably be gathered *[1]*.

Earlier studies *[2,3]* have shown that using the Global Positioning System (GPS) technology in travel surveys has the potential to enhance the quality and scope of travel data collection, without overburdening the survey participants. Earlier studies using GPS devices for travel data collection were usually done in conjunction with the use of personal digital assistants (PDA) *[4]*. Travel survey participants were required to enter additional trip information such as trip purpose and mode of travel. Such requirement for additional information not only represents an additional burden on survey participants, but it also increases the cost of travel surveys. A more efficient travel survey would rely mainly on data collected from a GPS device, without using PDA or any intervention while the participant is travelling. GPS travel data have to be interpreted accurately and efficiently in order to achieve the full range of benefits that the GPS technology can offer. Recent research efforts have been made to develop processing tools to identify trip ends, routes and activities from GPS travel data. However, most of these efforts have been focussed on a single mode of travel, which does not reflect the multimodal nature of travel behaviour. Any GPS-based travel survey conducted in an area that includes various travel modes will require a processing tool to identify the modes used by each survey respondent. Previous efforts made for mode identification of GPS travel data used simple rules relying mostly on travel speed thresholds *[5]*.

This project developed an integrated GPS-GIS system to automate the processing of GPS-based personal travel survey data. The project is divided into two main parts. The first involves a GPS-Alone System, which uses GPS travel data only as input. This system is intended for use in cases where GIS resources are not available. The second part involves developing a GPS/GIS Integrated System, which uses both GPS travel data and topological information in a GIS platform as inputs. This project is part of a multi-university project entitled “An Integrated GPS-GIS System for Collecting Spatio-temporal Microdata on Personal Travel in Urban Areas” which is funded by a Canadian Network of Centres of Excellence known as GEOIDE (Geomatics for Informed Decisions).

LITERATURE REVIEW

In the early stages of GPS deployment in travel surveys, the electronic travel diary (ETD) was used as a supplementary tool to GPS travel data collection. Samples of such studies include the Lexington study *[6]* and a multi-modal travel survey study done by Draijer et al. in the Netherlands *[7]*.

Because of the additional burden caused by the ETD, GPS has been experimentally used in travel surveys as a complete replacement of both traditional travel diaries (paper-and-pencil) and the ETD. In 1997, the first regional travel survey based on GPS and CATI was carried out in Austin, Texas *[8], [9]*. This study proved the feasibility of using GPS on vehicles for travel data collection without ETD. Activity dwell time, which is the minimum time duration for a stop between consecutive trips, was one of the important issues that early studies on GPS-based travel survey analysis attempted to address.

In Quebec City, Canada, a study using GPS to capture vehicle movements over a multi-day period was carried out to test a GIS-based algorithm which identifies route segments and locations of activities *[10]*. In another study by the Georgia Institute of Technology *[11]*, the process of data analysis involved trip detection, land use and address assignment, trip purpose identification, and travel route identification and distance calculation. A 120-second activity dwell time was applied for activity detection.

Other studies include a trip rate analysis using the California Statewide Household Travel Survey *[1]*, and a household travel survey in Sydney, Australia *[12]* which led to the development of customized applications including identification of trip ends, correction of cold starting problem and data losses within Urban Canyon areas, and presentation of colour-coded results on GIS maps.

A multi-modal travel survey involving a sub-sample of the 2002 London Area Transport Survey was also conducted *[13]*. The GPS data processing was performed by a Trip Identification and Analysis System developed at the GeoStats. A 120-second activity dwell time was applied. Mode identification was mainly based on the speed of GPS data. The study suggested that the nature of multi-modal trips and transfer time between modes required a more advanced algorithm for mode identification.

DESIGN OF GPS TRAVEL DATA ANALYSIS SYSTEM

While the GPS technology has a great potential to become a prime instrument for personal travel surveys, suitable analysis tools are required in order to extract the maximum travel information from the GPS data of each respondent. GIS tools best suit this analysis purpose. They provide accurate topology information for analysis and ideal visualization of trips and events. However, the cost of GIS software is expensive and a GIS transport network may not be available to every municipality, company or user. Besides, the data requirements for GIS are demanding. The level of detail and accuracy of the GIS road network model would affect the performance of GPS data analysis.

In order to design a GPS travel data analysis tool which is flexible enough for users with and without GIS transport networks and tools, two versions of an analysis system are developed in this research to automate GPS data analyses. The first version is a GPS-Alone System, which is independent of any GIS software and only GPS data uses for analyses. This system consists of four modules including data filtering, activity and trip identifications, multi-modal trip segment identification, and travel mode identification based on fuzzy logic. The Version 1 system is capable of identifying activities and trips, as well as the travel modes involved in each trip.

The second version of the analysis system is a GPS/GIS Integrated System, which employs GPS travel data, the output of the Version 1 system, and topological information from GIS. This system involves GIS-based link identification, as well as an interactive link matching-mode identification subsystem, which further refines the results from previous identifications performed separately in the Version 1 System. The interactive subsystem is comprised of five modules, including GIS mode identification, treatments of link matching failure, underground/indoor travel detection, off-road travel detection, and treatment of warm/cold starting problems. The analysis results can be presented to the survey participants during the prompted recall survey in order to confirm the identification results. The detailed survey design and technology used can be found in Tsui *[15]*. Figure 1 shows the structure of the GPS travel data analysis system.

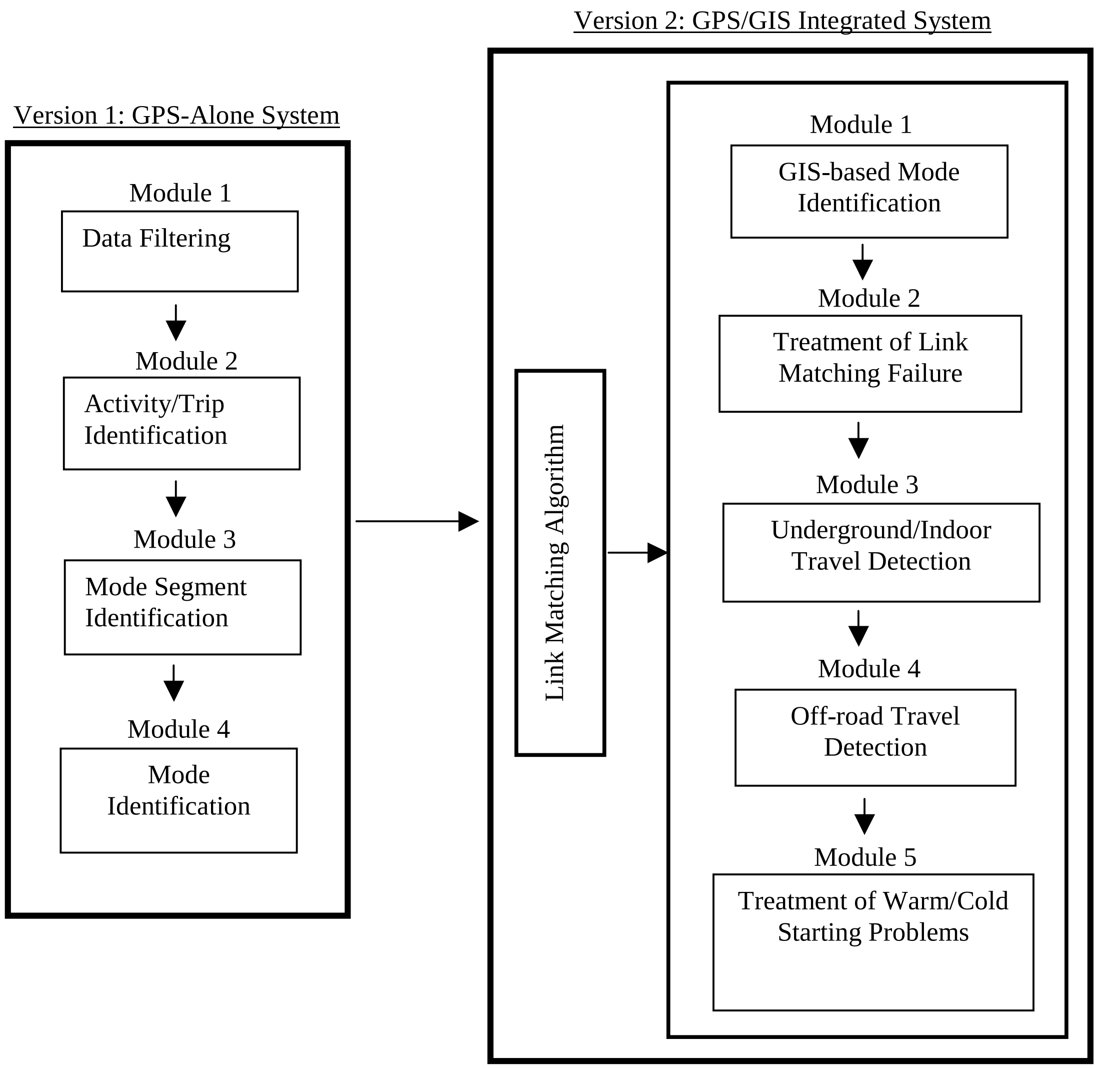

Figure 1: Structure of the GPS Travel Data Analysis System

VERSION 1 – GPS ALONE SYSTEM

Module 1 – Data Filtering

GPS data with poor quality are eliminated in this module by four data filters. The first two data filters are directly based on two parameters of each GPS record. GPS records with fewer than three satellites in view are eliminated. Other than the requirement for the number of satellites, GPS data records with Horizontal Dilution of Precision (HDOP) values higher than five (indicating low dispersion of satellites in view) are not considered in further analyses.

Although some GPS records satisfy the number of satellites and HDOP criteria, they do not provide correct information on heading and speed. Records carrying zero directional heading and zero speed are evidently erroneous when a GPS data trace is plotted on a map. Therefore, a third filter is applied to remove such GPS records.

The multipath error is a serious problem especially in "urban canyon" areas, causing GPS signals to jump around the area and form a data cloud instead of clear traces. These GPS data were removed by a forth filter. The original idea of this filter was developed by Chung and Shalaby *[5]* at the University of Toronto under the same

GEOIDE project and it has been embedded into this module. It removes GPS records which jump significantly from the original traces.

Module 2 – Activity and Trip Identification Algorithm

Module 2 is designed for activity and trip identification. Since each GPS raw data file includes records spanning one day, this module breaks down the one-day continuous GPS data records into individual activities and trips. A trip is defined as the connection between two activities. Identifying activities would thus identify trip ends.

The activity identification algorithm is a rule-based algorithm. Based on the literature and previous research, a 120-second dwell time was chosen as the basic criterion for activity identification *[13,14]*. There are two main categories in activity identification:

1. No GPS signal loss during an activity which may represent an outdoor activity, and
2. GPS signal loss during an activity.

The second category is sub-divided into long-duration indoor activities, short-duration indoor activities, and underground/indoor travelling activities.

Activity category 1 (no signal loss during an activity) was included for two purposes. First, when the survey participant is having an outdoor activity, the movement of the GPS antenna is minimal which yields a stream of GPS records with zero speeds. If the duration of such records is more than 120 seconds, this stream of GPS data records are grouped as one activity. Note that waiting time during transfer between transit routes maybe identified as such activity in some cases. There will usually be some small movements during an outdoor activity. Such movements will be dealt with by an activity merging action. The second purpose is for those participants who leave the GPS data logger in their car while undertaking activities close by. This kind of behaviour would also yield a stream of GPS points with zero speeds.

Activity category 2 (signal loss during activity) is sub-divided into three classes. The first class pertains to short-duration indoor activity. If the duration of a signal loss period ranges from 120 seconds to 600 seconds, and the distance travelled within the period is less than 50 meters, it is classified as a short-duration indoor activity. The lower bound of signal loss duration (120 seconds) is the minimum requirement for an activity dwell time, as suggested by literature *[13,14]*. Through empirical data analysis, it was found that 600-second is the most suitable threshold for the upper bound of signal loss duration for this class of activity. It is expected that for short-duration activities, the signal retrieval time when the participant comes out from indoors is fast enough such that the participant will not travel more than 50 meters from where he/she had the activity.

The second class of activity category 2 is intended for underground or indoor travelling activities. If the signal loss period is more than 120 seconds and the Euclidean distance travelled during this period is more than 500 meters, it is classified as an underground/indoor travelling activity. Since the spacing between subway stations is usually longer than 500 meters, any signal loss period with a Euclidean travel distance more than 500 meters may imply underground travel activity, which will be verified in Version 2.

The third class of activity category 2 is long-duration indoor activity. Any signal loss with duration more than 600 seconds, is classified as a long-duration activity. This kind of activity usually occurs when the participant is at home, work, school, etc.

Note that for activity category 2, there may be cases of signal loss duration between 120 and 600 seconds, and travel distance during the period ranging from 50 to 500 meters that do not fit into any of the above classes. Based on analysis of the GPS data, most of such cases are caused by signal loss while travelling by public transport on surface streets. There are frequent signal-loss periods while travelling on buses or streetcars, as the signal reception is often poor and not stable, particularly if the respondent is standing in the aisle. In the case of poor quality data which have frequent signal loss periods, the information contained in each sub-trip would be scarce. If such trip were separated into two, the information contained in the sub-trip might not be sufficient to support correct mode identification later in the analysis. Hence, the above classes of activity category 2 were designed in a way that prevents the erroneous identification of trips on buses or trams as activities, and as such the properties of the travel mode for the trip are retained.

Module 3 – Mode Segment Identification Algorithm

In the context of mid to large cities with multimodal transportation systems, many trips involve more than one mode. Therefore, it is crucial to separate each travel mode within a trip into segments, called mode segments, in order to differentiate properties of different travel modes. The point where travellers change from one travel mode to another is identified as a "mode transfer point" (MTP) in this project. It is used to separate two travel mode segments. Three kinds of potential MTPs are identified in the algorithm: end of walk (EOW) point, start of walk (SOW) point and end of gap (EOG) point. An EOW point is when a traveller changes from the walk mode to another travel mode, and *vice versa* for a SOW point. An EOG point is one after a long time gap which does not contain any GPS data. There is an implicit assumption in defining potential MTPs that walk is the intermediate travel mode in any mode transfer.

Potential MTP Identification Algorithm

The identification of EOW points is based on speed and acceleration of GPS data. When a transfer is made from walk to another mode, there is a significant difference in speed. As suggested by Chung and Shalaby [*5*], the boundary of when speed is changed from below 10 kph to above 10 kph is considered as an EOW point. In order to ensure a true speed-changing boundary is caused by mode transfer, the acceleration of GPS records are evaluated. Speed of walk is very stable throughout and acceleration during walking is very much close to zero. So, having the speed-changing boundary defined, the EOW point is found when three consecutive GPS records with acceleration smaller than 0.1 meter per second square are obtained, searching backward from the speed-changing boundary. As with the EOW point, all GPS records before this point should demonstrate walking property having speeds lower than 10 kph [5]. If any of the records violates this criterion, no EOW is defined for the speed-changing boundary. The identification of SOW points follows the same logic, except in the reverse order. An EOG point is defined when there is a time gap of more than 80 seconds which does not contain any GPS signal.

MTP Selection Algorithm

A set of potential MTPs (EOW, SOW, EOG points) for each trip is defined in the previous stage. In order to ensure reasonable and compatible MTP choices, a MTP selection algorithm was developed to choose MTPs from the potential points. The MTP selection algorithm is based on patterns and combinations of potential MTPs, speeds and duration of GPS records between potential MTPs. The selection process applies the following criteria:

- An EOW point should be followed by a SOW point and *vice versa*, if both of them are true MTPs. No consecutive EOW (or SOW) points are allowed to be determined as MTPs.
- An EOG point can either be EOW or SOW points.
- The segment of GPS records between an SOW and EOW point, identified as MTPs, should demonstrate walking properties. Speeds of all GPS points should be less than 10 kph and the duration of the segment should be longer than 60 seconds.
- The duration of a segment between an EOW and SOW points, identified as MTPs, should be longer than 120 seconds.

Note that transfer between 2 transit routes could also be identified by MTPs if there is a walking segment in between which satisfies the above criteria.

Module 4 – Fuzzy Logic Based Mode Identification Algorithm

Module 4 performs fuzzy logic-based mode identification for each mode segment, based on the characteristics of GPS records within the segment. Four characteristics of GPS records were chosen as the fuzzy variables for mode identification. They are:

- Average speed of GPS records (AS)
- 95th percentile maximum speed of GPS records (S)
- Positive median acceleration of GPS records (A)
- Data quality of GPS records (DQ)

The variable DQ is defined as the number of valid GPS records over the total GPS records in the segment. The actual values (crisp values) of the above fuzzy variables were found for each mode segment.

After selecting the fuzzy variables and obtaining their corresponding crisp values for each mode segment, the fuzzy inference system was developed. Further details can be found in Tsui *[15]*. Four classes of travel modes were considered in the fuzzy logic based mode identification. They are walk, cycle, bus and auto modes. A total of 17 rules were developed for the inference system.

- R1: If S is low and A is low then mode is WALK
- R2: If S is medium and A is low then mode is CYCLE
- R3: If S is high and A is low then mode is BUS
- R4: If S is low and A is medium then mode is WALK
- R5: If S is medium and A is medium then mode is CYCLE
- R6: If S is high and A is medium and AS is low then mode is BUS
- R7: If S is high and A is medium and AS is medium then mode is BUS
- R8: If S is high and A is medium and AS is high then mode is AUTO
- R9: If S is low and A is low then mode is BUS
- R10: If S is medium and A is high and AS is low then mode is BUS
- R11: If S is medium and A is high and AS is medium then mode is BUS
- R12: If S is medium and A is high and AS is high then mode is AUTO
- R13: If S is high and A is high and AS is low then mode is BUS
- R14: If S is high and A is high and AS is medium then mode is BUS
- R15: If S is high and A is high and AS is high then mode is AUTO
- R16: If S is high and A is high and DQ is bad then mode is BUS
- R17: If S is high and A is high and DQ is good then mode is BUS

With the help of the above rule base and a set of calibration data, the parameters for membership functions of the fuzzy variables were found by a neuro-fuzzy software called "NEFCLASS-J" developed by Nauck and Kruse *[16]*. Three triangular membership functions were used for the average speed, 95th percentile maximum speed and positive median acceleration, while two triangular functions were used for data quality.

Each travel mode class (e.g. Bus) contains more than one rule in the rule base. Each rule describes a characteristic of a class. A score will be assigned to each rule by the fuzzy inference system. Among the rules belonging to the same mode class, the rule with the maximum score best describes the mode segment. Thus, the likelihood of being a certain travel mode should be represented by the maximum score among the corresponding rules.

The membership functions with optimal parameters obtained from NEFCLASS-J, and the rule base were applied in the Module 4 algorithm. Based on the Set Theory, the AND operator favours the minimum membership values within a rule. The Max-Min Inference method was chosen because the rule with the maximum score best represents the mode segment. The fuzzy inference system would produce scores representing the possibility of travelling on each mode. The higher the scores of a travel mode class, the more possible that the mode was used by the traveller.

Distinct from other classification problems, the approach used here provides a partial membership to each travel mode class instead of assigning the mode segment to only one class. This is particularly useful when survey participants are asked to confirm their travel modes in the prompted recall survey. Survey participants can verify the results from a list of travel modes which should be ordered in accordance with the estimated likelihood of being the correct mode. The score of each class was expressed as a percentage of the total score of all classes. The percentage of being a certain travel mode is the likelihood that the travel mode was used in the mode segment.

VERSION 2 – GPS/GIS INTEGRATED SYSTEM

Link Matching Algorithm

The Link Matching Algorithm was developed by Chung and Shalaby *[5]* at the University of Toronto in 2003 under the same GEOIDE project. It is embedded in Version 2 GPS/GIS Integrated System as the main tool to identify travelled road links. There are three procedures in the algorithm: pre-processing, link matching and post-processing. Only the second and the third procedures are included in the Version 2 system.

The link matching procedure is the main matching process. It applies the approach proposed by Greenfeld *[17]* who used the coordinates and azimuth of GPS records for link matching. This procedure obtains preliminary matched links, and the final link selection is performed in the post-processing procedure.

Module 1 – GIS Mode Identification

This module performs a GIS mode identification, which aims at enhancing the fuzzy logic mode identification carried out in the Version 1 system. Apart from the filtered GPS records, and the mode identification results of Version 1, results from the Link Matching Algorithm and topology information from the GIS map are applied as well. The GIS mode identification was mainly designed to identify bus and streetcar travel modes. A GIS map with transit route service information would help to confirm if the transit service has been used by the survey participant in the mode segment.

A transit route searching algorithm forms the backbone of this module. The route searching algorithm will be activated when either the membership value for bus or cycle is greater than a threshold, which is set at 0.4. The membership value for cycle is also considered because the speed and acceleration characteristics of bicycles are similar to both streetcars and buses. When only the cycle membership value is greater than the threshold, the quality of the GPS records along the segment, defined as the number of valid records divided by the number of total records in the segment, has to be lower than a threshold set at 0.7. This extra criterion differentiates good-signal property while on a cycle from that on a bus or streetcar.

There are two procedures in the route searching algorithm: transit route searching for links and transit route selection. The first procedure performs a route search at a microscopic level, which checks the availability of routes for each selected link in the segment. The second procedure refines the route selection at a macroscopic level. It considers route availability for the mode segment as a whole. Transfers between routes have also been taken into consideration. Further details can be found in Tsui *[15]*.

After the route searching algorithm is performed, at most two travel modes will be assigned to mode segments. "Mode1" indicates the most possible travel mode while "Mode2" indicates the second most possible travel mode. The mode assignment scheme is shown in Table 1.

Table 1: Scheme for Mode Assignment

Cases				**Mode Assignment**			
Membership > 0.4			**DQ < 0.7**	**Transit Route found***		**No Transit Route found**	
Cycle	**Bus**	**Auto**		**Mode1**	**Mode2**	**Mode1**	**Mode2**
No	Yes	No	--	B/SC	--	A	--
No	Yes	Yes	--	B	A	A	--
Yes	Yes	No	--	B/SC	C	C	--
Yes	No	No	Yes	B/SC	C	C	--

Remarks:
- DQ: Data quality (No. of Valid Records / Total Records)
- A: Auto mode
- B/SC: Bus or streetcar mode
- C: cycle mode
- * Transit route is found along the matched links in the GIS map

For all other cases which do not activate the route searching algorithm, the two travel modes with the highest membership values would be assigned as "Mode1" and "Mode2" for the segment, given that these values are higher than the threshold value of 0.4.

Module 2 – Treatment of Link Matching Failure

There are cases where a continuous travel trajectory was not found and gaps appear between the final links selected by the Link Matching Algorithm. These situations are called link matching failures. Module 2 provides treatments of link matching failure cases, through filling the gaps between the final selected links. There are two approaches to treat the link matching failure: a transit route-based treatment and a link-based treatment, which is sub-divided into a treatment of poor GPS signals and a general treatment. The transit route-based treatment will be activated if at least one transit route is found for the mode segment in Module 1 (which implies transit service was used), otherwise the link-based treatment is carried out. The link-based treatment for poor GPS signals will be carried out for mode segments which do not contain any final selected link, while the segment with selected links will be taken care of by the general treatment of the link-based approach.

Transit Route-Based Treatment

The gap in between final selected links will be filled based on the information of the selected transit route of the mode segment. A link which contains the selected route of the segment will be chosen for the gap. The algorithm continues to search for links to fill in the gap until the gap is fully filled.

Link-Based Treatment – Poor GPS Signal

The mode segment is divided into sub-segments. Each sub-segment contains GPS records with heading direction of not more than 10 degrees difference from each other. The link which has the closest orientation with the heading direction will be chosen.

Link-Based Treatment – General

Each link is connected to two nodes and the gap between links is defined by a node from which the gap begins and a node at which the gap ends. They are called the entry and exit points. A set of candidate links which is connected to the entry point of the gap is then found. The selection of a link from these candidate links is based on one of the following cases. The sequence of cases shown below indicates the priority for link selection.

Case 1. If there is a link which connects the entry and exit points of a gap, it will be chosen for the gap filling. Otherwise, if there exists only one link in the candidate set of links, this link will be chosen for the gap filling.

Case 2. If the azimuth connecting the entry and exit points of the gap (called azimuth of the gap hereafter) is the same as either that of the link before the gap or the link after the gap, case 2 will be applied for gap filling. The link in the candidate set with the closest orientation as the azimuth of the gap will be chosen for the gap filling.

Case 3. If there are GPS records within the gap, the gap is divided into sub-segments, which contain GPS records with heading direction not more than 10 degrees of difference. The link in the candidate set which has the closest orientation (azimuth) as the heading direction is chosen.

Case 4. This case chooses the link in the candidate set that has the closest azimuth as the average azimuth of the links before the gap and the azimuth of the gap. There are two violations upon which this case will not be carried out. First, if the gap is greater than 1000 meters, case 4 will not be carried out since the combination of link choices for the gap increases with the gap distance. Second, if the orientation of the links before and after the gap is close to a right angle, case 4 will not be carried out since the traveller may make any right angle turn at any of the intersections.

Case 5. If there is any preliminary matched link in the candidate set, that was not selected by post-processing as a final selected link, this link will be chosen by Case 5.

Case 6. If none of the cases above fits for gap filling, case 6 will be carried out. An assumption was applied in this case that a traveller tends to avoid turning movements unless he is required to make a turn to arrive at the end of the gap. The link in the candidate set with the closest azimuth to that of the link before the gap will be chosen. If either nodes of this link lay beyond the link after the gap, this link is discarded and another link is chosen from the candidate set which has the closest azimuth as that of the link after the gap.

One link will be found at a time by one of the above cases. The algorithm iterates the procedure and a link is found for the gap based on the previous selected link, until the gap is fully filled. In order to prevent infinite link searching for the gap, if the number of links found for the gap is greater than a threshold and the gap is not fully filled, the gap filling will be terminated and all previous links found for the gap will be discarded. Further details and examples can be found in Tsui *[15]*.

Module 3 – Underground/Indoor Travel Detection

This module checks if there is underground or indoor travel when there is a signal loss period.

These underground or indoor trips were temporarily identified as an activity in Version 1. The last GPS record and the first GPS record before and after the activity will be checked to see if they are close to any subway station within a buffer of 0.004 decimal degree (which is 630 meters as defined in ArcGIS). If both GPS records are close to subway stations, a full membership of subway travel mode is assigned to the activity. If only one GPS record is close to a subway station, half subway membership is assigned. The selected subway station(s) is (are) highlighted on the GIS map for visualization. If both GPS records are not close to any subway station, subway travel is not considered and indoor travel may have been carried out or there was signal loss during travelling. If a segment has a full subway membership, the subway mode is assigned as the most probable travel mode ("Mode1"). If a segment has a half subway membership, subway is assigned as its "Mode1" while signal loss is assigned as its "Mode2". If a segment has a zero subway membership, signal loss will be assigned as its "Mode1".

Module 4 – Off-road Travel Detection

Off-road travel is checked by this module when the mode segment is a walking segment or the membership value for cycle is greater than 0.5. As GPS signal quality is usually very good while a traveller is walking or on a cycle, if there are at least 12 continuous GPS records which do not have a matched link, and that these records are following a steady heading direction, the algorithm would determine an off-road travel was involved.

Module 5 – Treatment of Warm/Cold Starting Problems

Module 5 targets the warm/cold starting problems and it tries to find links which fill in the gaps during these signal loss periods. Two main kinds of treatments are developed, one for the warm/cold starting problem after a traveller emerged from a subway station, and the other for signal loss period after an activity.

Warm/Cold Starting Treatment after Subway Travel

The closest node to the subway station from which the traveller exited is the entry point of the gap. The node connected to the first selected link of the next mode segment is the exit point of the gap. Links are found for the gap using the general link-based treatment described in Module 2, except that Case 2 and Case 4 are not performed because no link before the gap exists which is required by these two cases.

Warm/Cold Starting Treatment after an Activity

Warm/cold starting treatment will be activated if a gap exists between mode segments before and after an activity. The node of the last selected link in the segment prior to the activity is the entry point of the gap and the node of the first selected link in the segment following the activity is the exit point of the gap. The gap is filled with the general link-based treatment described in Module 2, except that Case 3 and Case 5 are not performed as there is no GPS records between the mode segments which is required by these cases.

TESTING AND ANALYSIS

The system discussed above was implemented using VBA (Visual Basic for Applications) and ArcObject on the ArcGIS platform. The system was tested using real GPS travel data collected with the GeoLogger™ (developed by GeoStats) in the City of Toronto, Ontario and some of its suburbs. A total of 9 volunteers have participated in the GPS travel data collection. Each participant was required to fill in a data log sheet describing his/her activities and trips. An instruction manual detailing the procedures for data collection and a sample data log sheet were given to the participants. In total, 58 days of GPS travel data were collected in 2004, including a total of 103 activities, 109 trips and 237 mode segments. Note that one GPS unit only was available for data collection, and therefore the 9 volunteers used it on separate days.

Activity Identification

All activities reported by participants (i.e. true activities) were correctly identified. The activity detection rate was 100%. Note that there were other activities detected that were not true activities. The "false" activities are mainly caused by cases that the traveller stops moving during a trip because of traffic conditions, yielding a series of zero-speed GPS records. Since the fuzzy logic based mode identification is based on the characteristics of travel modes, if these zero-speed GPS records were included in the trip, the average speed, 95^{th} percentile speed and the positive median acceleration of the travel mode would be lowered. Thus, the performance of travel modes would be affected by different traffic conditions and the characteristic measures of the travel mode would be distorted. In order to have true characteristic measures for each travel modes, these zero-speed GPS records were identified as false activities which eliminate the effect of traffic conditions. These "false" activities can be verified and eliminated by the survey participants during the prompted recall survey. Other "false" activities indicate the possibility of underground or indoor travelling activities, which will be detected and refined in the Version 2 system.

Mode Identification

Table 2 presents the mode identification results.

Table 2: Travel Modes Detection Rates - Version 1 and Version 2

	Version 1	Version 2	
	Average Detection Rate	Detection Rate	
		Type I	Type II
Walk	97%	98%	98%
Cycle	86%	72%	86%
Bus	76%	80%	80%
Auto	97%	97%	99%
Streetcar	--	88%	88%
Overall	91%	91%	94%

	Total Cases Reported by Participants	Cases Detected	
		True*	False#
Subway Travel	26	26	11
Off-Road Travel	43	42	6

Remarks:
Version 1: Fuzzy Logic Based Mode Identification
Version 2: GIS Mode Identification
Type I: Travel mode of a mode segment is correctly detected as Mode 1 (the most likely mode of travel)
Type II: Travel mode of a mode segment is correctly detected as Mode 2 (the second most likely mode of travel)
True*: True subway/off-road travel reported by participants
False#: False subway/off-road travel detected by algorithms

The overall detection rate of the Fuzzy Logic Mode Identification is 91%. Results of the walk and auto modes are as high as 97%, since speeds and accelerations of walk and auto are very distinguishable from other modes, and their data quality is usually very good. The result of the cycle mode is fairly good. The detection rate of the bus mode is only 76%, because some characteristics of bus performance overlap with those of other travel modes considerably. Besides, the large variability in bus performance also makes it hard to be identified.

Mode identification in Version 2 produced reasonable results, which are fairly comparable to that of Version 1. Mode identification for the walk and auto modes are good and consistent, ranging from 97% to 99%. Although Type I detection rate of the cycle mode is lower than that of Version 1, Type II detection rate of cycle mode is the same as Version 1. A lower Type I detection rate is mainly caused by cycle travelling along a bus/streetcar route. In such cases, cycle is assigned as the second most likely travel mode (Mode 2), while bus/streetcar travel mode is assigned as the most likely travel mode (Mode 1). Further mode verification would be required during the prompted recall survey.

The detection rate of bus increased from 76% in Version 1 to 80% in Version 2. It is enhanced by the transit route information provided on the GIS map. The improvement is not as high as expected due to poor data quality while on a bus/streetcar, which led in some cases to having no travelled links identified for some of the bus mode segments. As the algorithm requires at least one selected link in the mode segment in order to verify if a transit route has been used, no route could be found for a segment without any travelled links selected. Although there is not much improvement in the detection rate, the approach in Version 2 confirms the availability of transit routes while that of Version 1 is only based on characteristics of GPS records in the mode segment.

Streetcar was not included as one of the travel modes in Version 1 but it was included in Version 2, with a detection rate of 88%. Streetcar identification is effective using GIS topology information for streetcar mode segments. All subway travel was detected. The 11 false subway trips were due to signal loss near subway stations.

They can be eliminated during the prompted recall survey by the participants. The detection rate of off-road travel is very good, yielding 98%.

Link Identification

The GIS network model used in this study was for the City of Toronto. Trip data for 6 out of the available 58 days of data were not used for link identification because such data were for trips made outside Toronto. Table 3 shows the link identification results.

Table 3: Performance of GIS Link Identification

	Percentage of Correct Link Identification		Percentage Improvement[#]	
	Link Matching Algorithm (Chung, 2003)	Interactive Subsystem*		
Average	69%	94%	**Average Improvement**	25%
Maximum	100%	100%	**Maximum Improvement**	78%
Minimum	18%	54%	**Minimum Improvement**	0%

*Interactive Subsystem: Interactive Link Matching-Mode Identification Subsystem in the Version 2 system
[#]Percentage improvement by Interactive Subsystem, compared to Link Matching Algorithm

On average, 69% of the travelled links were found for each case by the Link Matching Algorithm and 94% of travelled links were found by the Interactive Subsystem. The link identification rate for the Link Matching Algorithm ranges from 18% to 100%, while that of the interactive subsystem ranges from 54% to 100%.

There are significant improvements in link identification by the Interactive Subsystem. For GPS travel data with good signal quality, both approaches yielding 100% identification rate. However, when the signal quality is bad, the identification rate goes down to 18% for the Link Matching Algorithm and 54% for the Interactive Subsystem. Despite of the poor signal quality, performance of the Interactive Subsystem exceeds that of the Link Matching Algorithm by 36% even in the worst case scenario. The average improvement is 25% and can be as high as 78%.

SUMMARY AND CONCLUSIONS

This project produced two versions of a system for GPS multimodal travel data analysis. Results from both versions were very promising. The GPS-alone System, which was developed using Visual Basic 6.0, allows activity and mode identifications independent of any GIS software. The GPS/GIS Integrated System provides link identification in a GIS platform as well as interactive link matching-mode identification algorithms to enhance the analysis.

This project demonstrates how GPS travel data analysis can be automated and highlights the benefits brought by an interactive analysis system, providing a new analysis method for personal-based GPS multimodal travel surveys. Further analyses and models could be built on top of these systems, such as the analysis for trip purpose and the activity scheduling models. In this dynamic 21st century of multimodal travel, the field of GPS travel survey research carries much exciting promise.

ACKNOWLEDGEMENTS

This research was supported by GEOIDE (Geomatics for Informed Decisions), a Canadian Network of Centres of Excellence.

REFERENCES

1. Wolf, J., Loechl, M., Myers, J. and Arce, C. Trip Rate Analysis in GPS-Enhanced Personal Travel Surveys. *International Conference on Transport Survey Quality and Innovation Kruger Park, South Africa*. August 2001.
2. Wagner, D., Murakami, E. and Neumeister, D. *Global Positioning System for Personal Travel Surveys*. FHWA, U.S. Department of Transportation, 1997.
3. Guensler, R., Wolf, J.. *Development of a Handheld Electronic Travel Diary for Monitoring Individual Trip Making Behaviour*. TRB, National Research Council, Washington, D.C., 1999.
4. Bachu, P. K., Dudala, T. and Kothuri, S. M. Prompted Recall in Global Positioning System Survey – Proof-of-Concept Study. *Transportation Research Record: Journal of the Transportation Research Board.* No. 1768, TRB, National Research Council, Washington, D.C., 2001, pp. 106-113.
5. Chung, E., and Shalaby, A. Development of a Trip Reconstruction Tool for GPS-Based Personal Travel Surveys, *Journal of Transportation Planning and Technology*, 2005, Vol. 28 No. 5 pp. 381-401.
6. Wagner, D. P. *Report: Lexington Area Travel Data Collection Test: GPS for Personal Travel Survey*. Final Report for OHIM, OTA and FHWA. Washington, D.C. Sept 1997.
7. Draijer, G., Kalfs, N. and Perdok, J. Global Positioning System as Data Collection Method for Travel Research. *Transportation Research Record: Journal of the Transportation Research Board*. No. 1719, TRB, National Research Council, Washington, D.C., 2000, pp. 147-153.
8. Pearson, D. Global Position System (GPS) and Travel Surveys: Results from the 1997 Austin Household Survey. *Paper presented at the 8th Conference on the Application of Transportation Planning Methods*, Corpus Christi, Texas, April 2001.
9. Wolf, J. *Using GPS Data Loggers to Replace Travel Diaries in the Collection of Travel Data*, Dissertation, Georgia Institute of Technology, School of Civil and Environmental Engineering, Atlanta, Georgia, July 2000.
10. Doherty, S. T., Noel, N., Lee-Gosselin, M. E. H., Sirois, C. and Ueno, M. Moving beyond Observed Outcomes: Integrating Global Positioning Systems and Interactive Computer-based Travel Behaviour Surveys. *Proceedings of the TRB Conference on Personal Travel: The Long and Short of It*. Washington, D.C.: Transportation Research Board, 1999.
11. Wolf, J., Guensler, R. and Bachman, W. Elimination of the Travel Survey Diary – Experiment to Derive Trip Purpose from Global Positioning System Travel Data. *Transportation Research Record: Journal of the Transportation Research Board*. No. 1768, TRB, National Research Council, Washington, D.C., 2001, pp. 125-134.
12. Stopher, P., Jiang, Q., and Bullock, P. *Making Sense of It All: Application for Managing and Analysing In-Vehicle GPS Data*. The University of Sydney, Australia, 2003.
13. Wolf, J. and Duckenfield, T. *The Use of GPS to Improve Travel Data – Use of GPS in Travel Surveys, Study Report*. London: Steer Davies Gleave, prepared for DTLR News Horizon Programme, 2003.
14. Stopher, P., Jiang, Q., and Bullock, P. *Making Sense of It All: Applications for Managing and Analysing In-Vehicle GPS Data*. Institute of Transport Studies. The University of Sydney. NSW 2006, Australia, 2003.
15. Tsui, S.Y.A., 2005. "An Enhanced System for Link and Mode Identifications for GPS-Based Personal Travel Surveys". Unpublished M.A.Sc. Thesis, Department of Civil Engineering, University of Toronto, 2005.
16. Nauck, D.. *NEFCLASS Neuro-Fuzzy Classification*. http://fuzzy.cs.uni-magdeburg.de/nefclass/nefclass.html. Accessed 16 November 2004.
17. Greenfeld, J. Matching GPS Observations to Locations on Digital Map. *81st Annual Meeting of the Transportation Research Board*, Washington, D.C., 2002.